Search for a heavy top $t' \Rightarrow Wq$ in top events

D. Cox, on behalf of the CDF Collaboration
Department of Physics, University of California-Davis, Davis, CA 95616, USA

We present a search for a massive quark (t') decaying to Wq and thus mimicking the top quark decay signature in data collected by the CDF II detector corresponding to 2.8 fb^{-1}. We use the reconstructed mass of the t' quark and the scalar sum of the transverse energies in the event to discriminate possible new physics from Standard Model processes, and set limits on a standard 4th generation t' quark.

1. Introduction

In this study we investigate whether present data from the CDF detector allow or preclude the production of a hypothetical new quark which decays to a final state with a high-p_T lepton, large $\not\!\!E_T$, and multiple hadronic jets having large total transverse energy E_T thus mimicking top quark pair event signatures. We refer to the hypothetical new quark as t' for brevity, although such a signature could be a standard fourth-generation up-type heavy quark or any up-type quark. A number of theoretical models advocate a fourth chiral generation of massive fermions with the same quantum numbers as ordinary ones [1, 2, 3, 4, 5]. Precision measurements from LEP exclude a light fourth neutrino ν_4 with mass $m(\nu_4) < m_Z/2$, where m_Z is the mass of the Z boson. A fourth generation neutrino cannot be too heavy due to sizeable radiative corections. Despite this, reasonable constructions of a fourth generation exist that remain viable [6]. In many theoretical models the addition of a fourth generation relaxes present bounds on the Higgs. In addition frequently a small mass splitting between new heavy quarks t' and b' is preferred, such that $m(b')+m(W) > m(t')$ and t' decays predominantly to Wq (a W boson and a down-type quark $q = d, s, b$) [7]. For the purposes of this analysis we assume that the new quark is pair-produced strongly, has mass greater than the top quark, and decays promptly to Wq final states. The data we use in the analysis corresponds to 2.8 fb^{-1} of integrated luminosity collected at the Fermilab Tevatron in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV.

2. Event Selection

The CDF II detector is described in detail in Ref. [8]. We can parameterize coordinates in the detector using the azimuthal angle ϕ and the pseudorapidity $\eta = -\ln[\tan(\theta/2)]$, where θ is the polar angle measured from the proton beam direction. The transverse energy E_T is defined as $E \sin\theta$, where E is the energy deposited in a calorimeter cluster. The transverse momentum p_T of a track is the component of the track momentum transverse to the beam-line. The missing transeverse energy $\not\!\!E_T$ is the magnitude of the vector defined as -$\Sigma_i E_T^i \hat{n}_T^i$, where $\hat{n}_T^i$ is the transverse component of the unit vector pointing from the interaction point to the calorimeter tower i. This is corected for the p_T of muons, which do not deposit all of their energy in the calorimeter, and tracks which point to uninstrumented regions in the calorimeter.

Highly energetic quarks (such as the down-type quark produced in the hypothesized t' decay) undergo fragmentation that results in jets of hadronic particles. Jet candidates are reconstructed using the calorimeter towers with corrections to improve the accuracy of the energy estimation and are required to have $\not\!\!E_T > 15$ GeV and $|\eta| < 2.5$.

Candidate events for this t' search are required to have exactly one isolated electron or muon with $p_T \geq 20$ GeV/c, $\not\!\!E_T \geq 35$ GeV, at least four jets with $E_T \geq 20$ GeV.

To reduce the contribution of the QCD background we also require a lead jet with $E_T \geq 60$ GeV and two cuts one in the $\Delta\phi$ between the corrected $\not\!\!E_T$ and the lepton vs $\not\!\!E_T$ plane (requiring $\Delta\phi \geq A_1 - (1/B_1)\not\!\!E_T$ where $A_1 = 4.408$; $B_1 = 6.11$) and one in the $\Delta\phi$ between $\not\!\!E_T$ and lead jet vs $\not\!\!E_T$ plane (requiring $\Delta\phi \geq A_2 - (1/B_2)\not\!\!E_T$ where $A_2 = 1.888$; $B_2 = 21.6$). The dominant contributing backgrounds after this event selection are from electroweak process as well as $t\bar{t}$ pair production. Electroweak processes are dominated by W + jets. For the $t\bar{t}$ background we use Monte Carlo (MC) generated using an assumed top quark mass of 175 GeV. The QCD background is modeled using a sample of data where the lepton ID cuts have been reversed. Other backgrounds (including Z+jets, WW+jets, WZ+jets and single top events) have a smaller rate than W + jets and in addition have been found to have similar kinematic distributions to W + jes and so are modeled as one background using the W + jets model.

3. Search Technique

We utilize the fact that in our regime of interest the t' decay chain is identical to that of the top quark to reconstruct the t' mass in the same way as is done in the top quark mass measurement analyses. We use the template method for top quark mass reconstruction [9]

based on the best χ^2-fit to the kinematic properties of the final top (or t') decay products. The χ^2 is given by the following expression:

$$\begin{aligned}\chi^2 &= \sum_{i=\ell,4jets} \frac{(p_T^{i,fit} - p_T^{i,meas})^2}{\sigma_i^2} \\ &+ \sum_{j=x,y} \frac{(p_j^{UE,fit} - p_j^{UE,meas})^2}{\sigma_j^2} \\ &+ \frac{(m_{jj} - m_W)^2}{\Gamma_W^2} + \frac{(m_{\ell\nu} - m_W)^2}{\Gamma_W^2} \\ &+ \frac{(m_{bjj} - m_t)^2}{\Gamma_t^2} + \frac{(m_{b\ell\nu} - m_t)^2}{\Gamma_t^2}\end{aligned} \quad (1)$$

where the invariant masses of the W decay products m_{jj} and $m_{\ell\nu}$ are constrained to the pole mass of the W boson, and the masses of top and anti-top (t' and $\bar{t'}$) quarks are required to be same. The jet and lepton energies as well as the unclustered energy (UE) are allowed to float within their resolution uncertainties. The transverse component of the neutrino momentum is determined as the negative sum of the lepton, jet and unclustered transverse energies:

$$\vec{p}_T^{\,\nu} = -(\vec{p}_T^{\,\ell} + \sum \vec{p}_T^{\,jet} + \vec{p}_T^{\,UE}). \quad (2)$$

For each event there are total $4!/2 = 12$ combinations of assigning 4 jets to partons. In addition, there are two solutions to account for the unknown z-component of the neutrino momentum. After minimization of the χ^2 expression, the combination with the lowest χ^2 is selected and the value of m_t is declared to be the reconstructed mass M_{reco} of top (or t').

We use the observed distributions of the M_{reco} and total transverse energy in the event

$$H_T = \sum_{jets} E_T + E_{T,\ell} + \not\!\!E_T \quad (3)$$

to distinguish the t' signal from the backgrounds by fitting it to a combination of t' signal, top, electroweak background, and QCD background shapes.

We use a binned in H_T and M_{reco} likelihood fit to extract the t' signal and/or set an upper limit on its production rate. We chose to use bins of 25 GeV in both H_T and M_reco with H_T in 26 bins from 150 to 800 GeV and M_reco in 16 bins from 100 to 500 GeV. The likelihood is defined as the product of the Poisson probabilities for observing n_i events in 2D bin i of (H_T, M_{reco}):

$$\mathcal{L}(\sigma_{t'}|n_i) = \prod_i P(n_i|\mu_i)$$

The expected number of events in each bin, μ_i, is given by the sum over all sources, indexed by j

$$\mu_i = \sum_j L_j \sigma_j \epsilon_{ij}$$

Here L_j is the integrated luminosity, σ_j is the cross section, and ϵ_{ij} is the efficiency per bin of (H_T, M_{reco}).

We calculate the likelihood as a function of the t' cross section, and use Bayes' Theorem to convert it into a posterior density in $\sigma_{t'}$. We can then use this posterior density to set an upper limit on (or if we get lucky, measure) the production rate of t'.

The production rate for W+jets is a free parameter in the fit. Other parameters are related to systematic errors and treated in the likelihood as nuisance parameters constrained within their expected (normal) distributions.

We adopt the profiling method [10] for dealing with these parameters, i.e. the likelihood is maximized with respect to the nuisance parameters. Taking this into account the likelihood takes the following expression:

$$\mathcal{L}(\sigma_{t'}|n_i) = \prod_{i,k} P(n_i|\mu_i) \times G(\nu_k|\tilde{\nu}_k, \sigma_{\nu_k})$$

where ν_k are the nuisance parameters, such as $\sigma_{t\bar{t}}$, L_j and etc. $\tilde{\nu}_k$ are their central nominal values and σ_{ν_k} are their uncertainties. G is a Gaussian function centered at $\tilde{\nu}_k$ of width σ_{ν_k}.

4. Systematic Errors

The sensitivity to t' depends on knowing accurately the distribution of (H_T, M_{reco}) in data. We consider several sources of uncertainty including the jet energy scale, W+jets Q^2 Scale, initial and final state radiation (ISR and FSR), Parton Distrubiton Function (PDF) uncertaintity and others.

4.1. Jet Energy Scale

The sensitivity to t' depends on knowing accurately the distribution of (H_T, M_{reco}) in data. One of the largest sources of uncertainty comes from a factor that has a large effect on the shape of the kinematic distribution, the jet energy scale. Jets in the data and MC are corrected for various effects as described in [11], leaving some residual uncertainty.

This uncertainty results in possible shifts in the H_T and M_{reco} distributions for both new physics and standard model templates. We take this effect into account by generating templates with energies of all jets shifted upwards by one standard deviation (+1 templates) and downwards (-1 templates) respectively.

We then use a template morphing technique that was developed in 2005 for a previous version of this analysis. We interpolate and extrapolate the expectation value μ_i at each bin i as follows:

$$\mu_i = \mu_{0,i} + \nu_{JES} \cdot (\mu_{+1,i} - \mu_{-1,i})/2 \qquad (4)$$

where $\mu_{0,i}$ is the nominal expectation value, $\mu_{-1,i}$ and $\mu_{+1,i}$ are the expectation values from (-1) and (+1) templates respectively,andν_{JES} is the nuisance parameter representing the relative shift in jet energy scale:

$$\nu_{JES} = \frac{\Delta_{JES}}{\sigma_{JES}} \qquad (5)$$

It enters the likelihood (4) as a Gaussian constraint penalty term: $G(\nu_{JES}|0,1) = \frac{1}{\sqrt{2\pi}} e^{-\nu^2_{JES}/2}$.

4.2. W+jets Q^2 Scale

The effect of the choice of the appropriate Q^2 scale for W+jets production is evaluated by measuring the resulting change in the measured t' cross section given that t' exists. The Q^2 scale is varied to $2Q^2_{nominal}$ and $\frac{1}{2}Q^2_{nominal}$ the expected change in the measured cross section is then interpreted as the uncertainty on the t' cross section itself. We measure this shift as a function of the t' cross-section by drawing pseudoexperiments from shifted templates and fitting them to the nominal distribution. The resulting shift is fitted to a linear function of that t' cross-section and is incorporated into the likelihood as an additive parameter to the t' cross section, so that the t' contribution to the expectation value μ_i (4) in bin i becomes

$$\mu_{i,t'} = L_{t'}(\sigma_{t'} + \nu_{Q^2})\epsilon_{i,t'} \qquad (6)$$

where ν_{Q^2} is constrained by a gaussian with a width, that is a half of the largest of the upwards or downwards shifts for each mass of the t'.

The Q^2 systematic uncertainty for the different t' masses are shown in Table I.

$m(t')$ (GeV)	Systematic Uncertaintity (pb)
180	0.065
200	0.044
220	0.021
240	0.011
260	0.013
280	0.009
300	0.007
320	0.005
340	0.006
360	0.005
380	0.003
400	0.003
450	0.003
500	0.002

Table I Q^2 Systematic error for each t' mass point.

4.3. ISR and FSR

We varied the amount of initial- and final-state radiation together, i.e. shifting both up or both down. We generated samples with more ISR and FSR as well as some with less ISR and FSR. We refer to these samples as IFSR more and IFSR less. We generated samples for t' with masses of 250, 300 and 350 GeV which brackets the region where we expect to be able to place our exclusion limit.

The resulting effect is treated in a similar way to the Q^2 systematic. Templates are made for each of these mass points. Pseudoexperiments are thrown with the shifted top and t' IFSR samples, where the shift is set to be the same for top and t'. We then fit the obtained cross-section shift using a linear function of the t' cross-section.

The resulting shifts are shown in Table II. We add the resulting shifts in quadrature with the Q^2 error in the likelihood.

$m(t')$ (GeV)	IFSR offset	IFSR slope
180	0.125	0.026
200	0.125	0.024
220	0.125	0.022
240	0.110	0.020
260	0.080	0.018
280	0.060	0.017
300	0.035	0.014
320	0.025	0.011
340	0.015	0.009
360	0.010	0.008
380	0.007	0.007
400	0.005	0.006
450	0.004	0.005
500	0.003	0.004

Table II Columns 2+3: Shift (in picobarns) in apparent t' cross section due to a shift in the initial- and final-state radiation up or down

4.4. QCD Background

The QCD background shape is modeled from a sample of data in which the electron cuts have been reversed. The QCD normalization is obtained by fitting

the background (electroweak, top, and QCD) distrubtions to the data with the missing E_T cut removed and then computing how much remains after all cuts are applied. As most of the QCD is expected to be found at low missing E_T.

We investigated using leptons that fail the isolation requirements to model our QCD background. This gave an excellent description of the jet E_T spectra but a very poor description of the lepton p_T distributions, which were much steeper in this model than expected from the data. On the other hand the sample requiring the opposite of the electron cuts seems to give a reasonable description of most of the data kinematic distributions but has only very limited statistics; this means that the QCD templates need to be scaled up. Cutting very hard on the leading jet E_T removes most of the QCD background which makes our fit rather insensitive to the QCD modeling. The relative normalization uncertainty is taken to be 50%, due to our lack of confidence in our model and normalization method. With our QCD veto cuts it turns out to change the fit by a negligible amount whether we constrain QCD or let it float. The uncertainty is represented by a Gaussian-constrained parameter in the likelihood. The QCD background has a negligible effect on the t' limit.

4.5. Integrated Luminosity

The integrated luminosity uncertainty is taken to be 5.9%, and is represented by an additional gaussian-constrained parameter multiplying all contributions except for the QCD background, which is normalized from data.

4.6. Lepton ID

We have two components for lepton ID. First is the efficiencies for the individual electrons and muons. We multiply each lepton type by the associated efficiency and gaussian constrain it within the error on the efficiency.

Second is the uncertainty on the lepton ID efficiency data/MC scale factor, which is of 2%, and taken as correlated across lepton types since it is due to the presences of multiple jets in an event. We add it in quadrature with the luminosity error, which is also correlated across lepton types, and include it with a gaussian constraint into the likelihood.

4.7. PDF Uncertainty

The PDFs are not precisely known, and this uncertainty leads to a corresponding uncertainty in the predicted cross sections, as well as the acceptance.

This effect is evaluated on both the top and the t' MC samples. The method consists in re-weighting the existing MC samples by the relative PDF weights given the parton momentum fractions (x_1, x_2) and Q^2 of the generated interaction.

46 eigenvectors are considered. We look at the difference between pairs of the CTEQ6M PDFs and add up these in quadrature. We then consider the difference between the two MRST72 and CTEQ5L PDF sets. If this is smaller than the 20 PDF sets uncertainty, we drop it. If it is larger, we add it in quadrature. To investigate the effect of α_s we look at the difference between the MRST72 and MRST75 PDF sets and add this in quadrature to the above errors.

The final PDF uncertaintes are given for each t' mass point as well as for top in Table III. A common conservative systematic error is added in quadrature to all other multiplicative factors and it taken as 1.1% for all templates.

	top	
175	+0.0110	-0.0112
	tprime	
180	+0.007	-0.008
200	+0.004	-0.005
220	+0.005	-0.005
240	+0.003	-0.003
260	+0.003	-0.003
280	+0.002	-0.003
300	+0.001	-0.003
320	+0.001	-0.002
340	+0.002	-0.002
360	+0.003	-0.002
380	+0.002	-0.002
400	+0.005	-0.002
450	+0.004	-0.005
500	+0.015	-0.013

Table III PDF uncertainty on top and t' calculated by reweighting the events according to the probability (given the various PDFs) of finding an up and down quark with appropriate momentum fractions.

4.8. Theory Uncertainty

The theory uncertainty in the t' cross section is about 10% (see Table IV), which is mainly due to uncertainty in PDFs ($\sim$ 7%). The other effect comes from uncertainty in the choice of the Q^2 scale [12].

We take the theory uncertainty in $t\bar{t}$ cross section fully correlated with the one of $t'\bar{t}'$, and introduce it into the likelihood as a single nuisance parameter: $\nu_{theory} = \nu_{theory}(m'_t)$.

5. Results and Conclusion

We tested the sensitivity of our method by drawing pseudoexperiments from standard model distributions, i.e. assuming no t' contribution. The ranges of the expected 95% CL upper limits with one and two standard deviation bandwidth are shown in Figure 1. The purple curve is the theoretical prediction [12, 13],

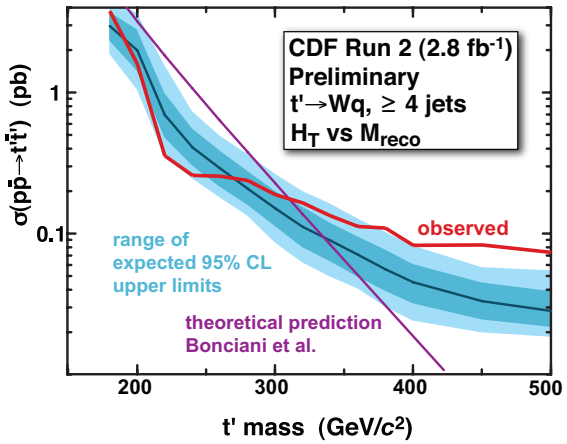

Figure 1: Upper limit, at 95% CL, on the production rate for t' as a function of t' mass (red). The purple curve is a theoretical cross section. The dark blue band is the range of expected 95% CL upper limits within one standard deviation. The light blue band represents two standard deviations

the values of which are given in Table IV. The lower σ_{min} and upper σ_{max} limits are obtained using the CTEQ6M family of parton density functions with uncertainties, together with the study of the scale uncertainty [14]. From Figure 1 it follows that given no t'

$m(t')$ (GeV)	σ_{min} (pb)	σ_{center} (pb)	σ_{max} (pb)
180.0	4.9938	5.7476	6.2396
200.0	2.7815	3.1898	3.4525
220.0	1.5926	1.8236	1.9710
240.0	0.9299	1.0647	1.1515
260.0	0.5499	0.6302	0.6828
280.0	0.3281	0.3769	0.4096
300.0	0.1968	0.2268	0.2475
320.0	0.1183	0.1370	0.1502
340.0	0.0711	0.0828	0.0914
360.0	0.0426	0.0500	0.0555
380.0	0.0255	0.0301	0.0337
400.0	0.0152	0.0181	0.0204

Table IV Theory values of t' cross section for given mass [12, 13].

presence, this method is on average sensitive to setting an upper limit at 311 GeV t' mass. The red curve in Figure 1 shows the final result expressed as a 95% CL upper limit on the t' production rate as a function of t' mass. Table V shows the individual calculated limits along with expected limits from pseudo-experiments for reference.

Based on these results we exclude at 95% CL a t' quark with a mass below 311 GeV, given that the true top mass is 175 GeV. Of course, our measurement of the top mass may have been affected by the presence of a higher mass t' and thus we should treate these conclucsions with care.

The 2D-distriubtion of (H_T, M_{reco}) is shown in Figure 3. Distributions of H_T and M_{reco} showing the result of the fit for $m(t')$=300 GeV are shown in Figure 2

To determine if the data show any evidence of an excess in the tails of H_T and M_{reco}, we decided *a priori* to count the number of events in groups of $n \times n$ of our standard 25 GeV bins in these quantties, and compare with the number predicted from a zero-signal fit to the full two dimensional spectrum. For each $n \times n$ bin one can then calculate the p-value for having observed that number or greater, given the prediction. If a significant effect is observed, one can calculate an overall p-value which is the probability that one would observe a p-value at least as significant as the most significant $n \times n$ bin or greater; this takes into account both the trials factor and the effect of systematic errors. Table VI shows the result of this counting experiment. The most significant $n \times n$ bin is for $n = 10$; the probability for observing 29 or more events given 18.03 expected is 0.01. (This assumes systematic uncertainty on the background.) Thus we conclude that there is no statistically significant excess in the far tails of H_T and M_{reco}.

$m(t')$ (GeV)	expected limit (pb)	observed limit (pb)
180	$2.954\ ^{+0.818}_{-0.592}$	3.759
200	$1.959\ ^{+0.869}_{-0.525}$	1.595
220	$0.693\ ^{+0.309}_{-0.207}$	0.355
240	$0.406\ ^{+0.152}_{-0.100}$	0.258
260	$0.288\ ^{+0.102}_{-0.072}$	0.254
280	$0.208\ ^{+0.072}_{-0.047}$	0.237
300	$0.150\ ^{+0.052}_{-0.034}$	0.188
320	$0.112\ ^{+0.044}_{-0.031}$	0.165
340	$0.088\ ^{+0.032}_{-0.024}$	0.133
360	$0.070\ ^{+0.026}_{-0.020}$	0.112
380	$0.056\ ^{+0.021}_{-0.016}$	0.109
400	$0.045\ ^{+0.017}_{-0.013}$	0.081
450	$0.033\ ^{+0.012}_{-0.008}$	0.083
500	$0.028\ ^{+0.011}_{-0.006}$	0.073

Table V Expected and obtained limits on t' production cross section for given mass.

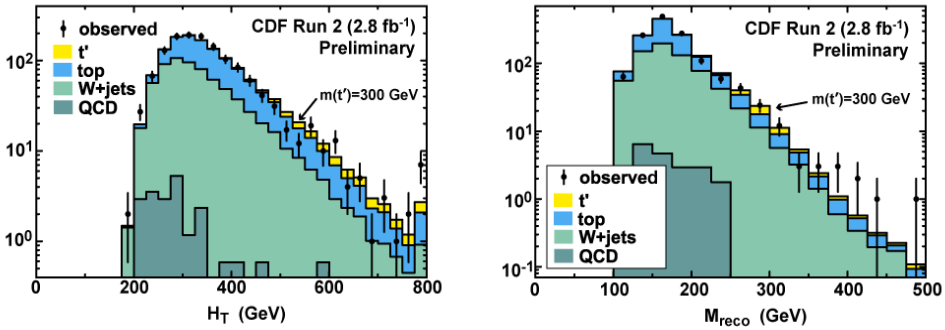

Figure 2: H_T (left & M_{reco} (right) distriubtions showing the results of the fit for m(t') = 300 GeV. The normalizations of the various sources and distortions of the kinematic distributions due to systematic effects are those corresponding to the maximum likelihood when the cross section for t' is set to its 95% CL upper limit.

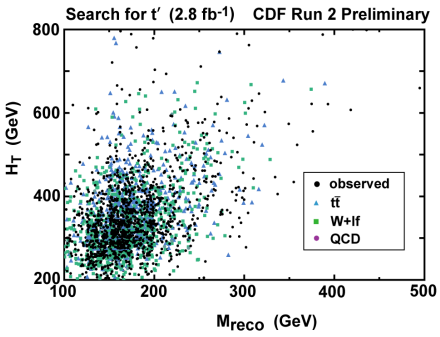

Figure 3: 2D distriubtion of H_T vs M_{reco} distribution showing hte data (black points) and the fitted number of background events: QCD (dark cyan circles), W + jets (green squares), and $t\bar{t}$ (blue triangles).

n	Min M_{rec} [GeV/c^2]	Min H_T [GeV]	observed	expected	p-value
1	475	775	0	0.021	1.000
2	450	750	0	0.116	1.000
3	425	725	1	0.228	0.2040
4	400	700	2	0.371	0.0540
5	375	675	3	0.718	0.0364
6	350	650	4	1.503	0.0660
7	325	625	4	2.876	0.3251
8	300	600	12	5.498	0.0110
9	275	575	14	9.885	0.1273
10	250	550	29	18.03	0.0105
11	225	525	41	31.34	0.0555
12	200	500	58	52.05	0.2219
13	175	475	92	91.14	0.4779
14	150	450	152	158.7	0.7141
15	125	425	222	231.0	0.7318

Table VI Number of observed events in the highest $n \times n$ bins of H_T and M_{reco}, compared with the prediction from a zero-signal fit to the full spectrum. For each value of n the table shows the p-value, the probability for observing at least what was actually observed or more, given the number expected. The minimum H_T and M_{reco} values in each trial are also shown.

References

[1] J. Silva-Marcos JHEP 0212 (2002) 036, [arXiv:hep-ph/0204217]; S. Sultansoy *et al.* , Acta Phys.Polon. **B37** (2006) 2839-2850, [arXiv:hep-ph/0502050]

[2] N. Borstnik *et al.*, Bled workshops in physics, Vol.7, No. 2, DMFA-Zaloznistvo, Ljubljana, Dec. 2006, [arXiv:hep-ph/0612250].

[3] T. Han, *et al.*, Phys. Lett. B563:191 (2003).

[4] H.-J. He, N. Polonsky and S. Su, [arXiv:hep-ph/0102144]

[5] C. Wagner *et al.*, [arXiv:hep-ph/0109097]

[6] Graham D.Kribs, Tilman Plhen, Michael Spannowsk and Tim M.P. Tait, [arxiv:hep-ph/0706.3718]

[7] P.Frampton, P. Hung, and M. Sher, Phys. Rept. 330, 263 (200).

[8] A. Abulenia *et al.* (CDF Collaboration), J. Phys. G 34, 2457 (2007).

[9] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. D **73**, 032003 (2006)[arXiv:hep-ex/0510048].

[10] J. Conway *et al.*, CDF Public note 7113.

[11] A. Bhatti *et al.*, Submitted to Nucl. Instr. Meth. A [arXiv:hep-ex/0510047].

[12] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason and G. Ridolfi, JHEP **0404** (2004) 068 [arXiv:hep-ph/0303085].

[13] R. Bonciani, S. Catani, M. L. Mangano and P. Nason, Nucl. Phys. B **529** (1998) 424 [arXiv:hep-ph/9801375].

[14] Personal communication with Michelangelo Mangano.